

**BY ORDER OF THE SECRETARY
OF THE AIR FORCE**

AIR FORCE INSTRUCTION 15-114

16 MARCH 2017



Weather

**WEATHER TECHNICAL
READINESS EVALUATION**

COMPLIANCE WITH THIS PUBLICATION IS MANDATORY

ACCESSIBILITY: Publications and forms are available on the e-Publishing website at www.e-Publishing.af.mil for downloading or ordering

RELEASABILITY: There are no releasability restrictions on this publication

OPR: HQ USAF/A3WP

Certified by: HQ USAF/A3W
(Mr. Ralph O. Stoffler)

Supersedes: AFI 15-114,
7 December 2001

Pages: 37

This Air Force (AF) Instruction (AFI) implements AF Policy Directive (AFPD) 15-1, *Weather Operations* and provides weather personnel and their organizational commander's guidance on how to evaluate products based on quantifiable measures of operational performance. This instruction applies to all organizations in the United States Air Force (USAF) with weather personnel assigned, to include Air Force Reserve Command (AFRC) and Air National Guard (ANG) and will be incorporated into government-contracted weather operations Statement of Work or Performance Work Statement. **(T-1)**. This AFI may be supplemented at any level, but all supplements that directly implement this publication must be routed to the office of primary responsibility (OPR) for coordination prior to certification and approval. **(T1)**. Refer recommended changes to the OPR using AF Form 847, Recommendation for Change of Publication; route AF Form 847s from the field through the appropriate functional office within the chain of command. The authorities to waive wing/organization level requirements in this publication are identified with a Tier number ("**T-0, T-1, T-2, T-3**") following the compliance statement. See AFI 33-360, *Publications and Forms Management*, for a description of the authorities associated with the Tier numbers. Submit requests for waivers through the chain of command to the appropriate Tier waiver approval authority. For non-tiered and Tier 1 compliance items, HQ USAF/A3W is the waiver approval authority for this publication; submit waiver requests to the OPR for processing. Ensure that all records created as a result of processes prescribed in this publication are maintained in accordance with AF Manual (AFMAN) 33-363, *Management of Records*, and disposed of in accordance with the AF Records Disposition Schedule located in the AF Records Information Management System. Elements of this instruction that require modification of existing software become effective 1 year after the date of this publication. Units will continue collecting and reporting metrics following their

established procedures until software modifications are made or when this publication becomes effective, whichever occurs first. Where capability exists, units implement this instruction immediately.

SUMMARY OF CHANGES

This document was completely revised and must be thoroughly reviewed. It now contains new detailed roles and responsibilities and directs weather units from the MAJCOM down to the flight and detachment level. New technical readiness metrics and measures of performance are provided in detail along with the units responsible for reporting them to their parent chain of command.

Chapter 1— WEATHER TECHNICAL READINESS PROGRAM OVERVIEW	4
1.1. Overview.....	4
Chapter 2— ROLES AND RESPONSIBILITIES	6
2.1. The AF Director of Weather (HQ USAF/A3W) will:	6
2.2. MAJCOMs will:	6
2.3. Organizations that produce forecast WWAs.....	7
2.4. Organizations that produce TAFs will:	7
2.5. Organizations that produce WPs will:	7
2.6. Organizations that produce MEFs will:	8
2.7. Organizations that produce Graphical Weather Depictions will:	8
2.8. Organizations that produce Numerical Weather Model Depictions will:	9
Chapter 3— WEATHER WATCH, WARNING AND ADVISORY VERIFICATION	10
3.1. WARNVER Guidance and Procedures.	10
3.2. WARNVER Statistical Evaluation Methods	10
Table 3.1. WARNVER MOPs.	11
Table 3.2. WARNVER Technical Readiness MOPs.	12
Table 3.3. WARNVER Standards.	13
Chapter 4— TERMINAL AERODROME FORECAST VERIFICATION	14
4.1. TAFVER Guidance and Procedures.	14
4.2. TAFVER Statistical Evaluation Methods and MOPs.....	14
Table 4.1. TAFVER MOPs.....	15

Table 4.2.	TAFVER Technical Readiness Metrics.....	18
Chapter 5—	OPERATIONAL VERIFICATION	19
5.1.	OPVER Guidance and Procedures.....	19
5.2.	OPVER Scope.....	19
5.3.	OPVER Statistical Evaluation Methods and MOPs.	19
Table 5.1.	WP/MEF Grid.....	21
Table 5.2.	WP/MEF Metrics.	22
Table 5.3.	WP/MEF Technical Health Metrics.....	23
Chapter 6—	GRAPHICAL WEATHER DEPICTION VERIFICATION	24
6.1.	GRAPHVER Guidance and Procedures.	24
6.2.	GRAPHVER is based on observed conditions or derived evaluations based on reliable information throughout the valid period of the graphics product in addition to subjective verification data points.	24
6.3.	Objectively grade turbulence and icing product sets.	24
6.4.	Objectively grade the thunderstorm product sets.....	25
Chapter 7—	NUMERICAL WEATHER MODEL VERIFICATION (MODVER)	27
7.1.	MODVER Guidance and Procedures.	27
7.2.	Statistical Evaluation Methods and MOPs	27
Table 7.1.	Definition of Variables.	29
Table 7.2.	Required Minimum MOPs.....	30
Table 7.3.	Optional MOPs for Determinist NWM Output.....	31
Table 7.4.	Optional MOPs for Stochastic NWM Output.....	32
Attachment 1—	GLOSSARY OF REFERENCES AND SUPPORTING INFORMATION	35

Chapter 1

WEATHER TECHNICAL READINESS PROGRAM OVERVIEW

1.1. Overview . The weather technical readiness program measures performance of weather functional capabilities (characterize and exploit weather information) and processes in support of AF Service Core Functions. Metrics measure operational performance of meeting end-user requirements, determine trends and provide analysis data for supported organizations and senior leaders. The weather metrics program focuses on measuring performance to provide timely, relevant, accurate and consistent environmental information to decision makers and commanders at all levels and to direct resources toward improvement areas. Commanders should use these metrics as one criterion to assess their weather readiness when evaluating their ability to meet Mission Essential Tasks in Defense Readiness Reporting System. Weather metrics include:

1.1.1. Weather Watch, Warning and Advisory (WWA) Verification (WARNVER): A threshold-based measure of performance (MOP) that reports WWA accuracy and timeliness by measuring whether the criteria stated in the WWA were met or not met according to the predetermined desired lead time (DLT). [Chapter 3](#) details specific WARNVER processes, procedures, MOPs, technical readiness metrics and standards.

1.1.2. Terminal Aerodrome Forecast (TAF) Verification (TAFVER): A threshold-based product that reports TAF accuracy by verifying forecast conditions against observed conditions including specific mission-critical weather phenomena and thresholds. [Chapter 4](#) details TAFVER responsibilities, MOPs, technical readiness metrics and standards.

1.1.3. Operational Verification (OPVER): A threshold-based metric that reports when and how forecast weather information impacts mission planning and when actual weather phenomena impact mission execution. OPVER measures the performance of Weather Products (WPs) and Mission Execution Forecasts (MEFs). [Chapter 5](#) details specific OPVER processes, MOPs, technical readiness metrics and standards.

1.1.3.1. WPs are defined as products generated by weather personnel for weather personnel or for non-weather personnel to use for planning purposes, situational awareness, and MEF generation. WPs include but are not limited to: military operations area forecasts (MOAFs), TAFs, air refueling forecasts, air combat maneuver/training area forecasts, instrument flight rules (IFR) military training route forecasts, drop/landing zone (DZ/LZ) forecasts, training range forecasts, and control forecasts for an operation with multiple missions. Although a TAF is a WP, TAFVER is its own program outside of OPVER.

1.1.3.2. MEFs are defined as products generated by weather organizations that are focused on execution of aviation, space, ground (Operations Group, Mission Support Group or other organizational operations) or maritime operations. MEFs include but are not limited to: Department of Defense (DD) Form 175-1, *Flight Weather Brief*, verbal forecasts, computer-based presentation briefing software, flimsies for local flying, and other non-standard forms that are given to an operator for mission execution.

1.1.4. Graphical Weather Depiction Verification (GRAPHVER): A threshold-based metric that reports graphical weather depiction accuracy by verifying forecast depictions against

reported conditions and other subjective verification data points. **Chapter 6** details the procedures and requirements for GRAPHVER.

1.1.5. Weather Model Prediction Verification (MODVER): A metric that reports weather model accuracy by verifying model forecasts against observed weather conditions. **Chapter 7** details the procedures and requirements for MODVER.

Chapter 2

ROLES AND RESPONSIBILITIES

2.1. The AF Director of Weather (HQ USAF/A3W) will:

- 2.1.1. Use metrics to assist in evaluating the overall technical readiness of the weather functional community.
- 2.1.2. Direct policy and training changes to improve the weather functional community's technical performance in supporting AF, Army, DoD and Joint operations.
- 2.1.3. Provide template spreadsheets to assist metrics calculations and data organization for MAJCOMS.
- 2.1.4. Upon request, provide and/or arrange scientific services and technical assistance to MAJCOM weather functional staffs and weather field organizations.
- 2.1.5. Review and analyze WARNVER, TAFVER, OPVER, GRAPHVER and MODVER data provided by MAJCOMS. Provide MAJCOMS feedback relating to identified issues to include best practices and improvement opportunities.
- 2.1.6. Provide, and/or oversee development and implementation of automated capabilities for weather metrics.

2.2. MAJCOMs will:

- 2.2.1. Collect and consolidate WARNVER, TAFVER, OPVER, GRAPHVER and MODVER metrics data for weather organizations within their command. **(T-1).**
- 2.2.2. Use WARNVER, TAFVER, OPVER, GRAPHVER and MODVER metrics to monitor technical performance of weather operations within their commands. **(T-1).**
- 2.2.3. Provide command-specific guidance to subordinate organizations regarding metrics. MAJCOMs may direct additional MOPs, technical readiness metrics, and standards for MAJCOM unique requirements or mission considerations in supplements to this publication. **(T1).**
- 2.2.4. Provide guidance to subordinate organizations for reporting, analysis and exploitation of metrics data. **(T-1).**
- 2.2.5. Consolidate monthly data into the HQ USAF/A3W approved template for WARNVER, TAFVER, OPVER, GRAPHVER, and MODVER. Data is due to HQ USAF/A3W no later than (NLT) the 28th of the month for the previous month. E-mail the data to HQ USAF/A3W Workflow mailbox (usaf.pentagon.af-a3-5.mbx.a3w-weather-workflow@mail.mil) with a courtesy copy to the HQ USAF/A3WP Workflow mailbox (usaf.pentagon.af-a3.mbx.a3wp-weather-policy-workflow@mail.mil). MAJCOMS may request HQ USAF/A3W work directly with subordinate units to collect data if they are resource constrained. **(T-1).**
- 2.2.6. Submit recommendations for improved verification methods or tools developed by organizations under their command to HQ USAF/A3W for consideration as benchmarks and inclusion into policy. **(T-1).**

2.3. Organizations that produce forecast WWAs will:

- 2.3.1. Establish and maintain a WARNVER program to assess WWA performance, analyze trends and identify/address forecast technique and/or training shortfalls as required. (T-1).
- 2.3.2. Collect and report WARNVER MOPs for supported locations according to [Chapter 3](#). (T-1).
- 2.3.3. Assess operational performance using WARNVER MOPs, identify and document performance trends at the organization level. In addition, organizations will assess performance of individual weather personnel, identify improvement areas and direct performance improvement measures or additional training as required. (T-2).
- 2.3.4. Provide WARNVER metrics for all supported locations to their parent MAJCOM and, upon request, to their supported unit commanders. (T-1).
- 2.3.5. Create a monthly report that identifies WWA performance shortfalls and corrective actions taken. Include areas of exceptional performance so leadership can cross feed them to other organizations. Send these reports to their parent MAJCOM and, upon request, to their supported unit commanders. (T-1).

2.4. Organizations that produce TAFs will:

- 2.4.1. Establish and maintain a TAFVER program to assess TAF performance, analyze trends and identify/address forecast technique and/or training shortfalls as required. (T-1).
- 2.4.2. Collect and report TAFVER MOPs for all supported locations according to [Chapter 4](#) for both the model-generated (no Forecaster-in-the-Loop [FITL]) and the final FITL TAFs). (T-1). During backup operations model-generated TAFs will not be available and will not be verified; FITL TAFs created during backup will be verified according to [Chapter 4](#). (T-1).
- 2.4.3. Assess performance using TAFVER metrics and document performance trends at the organization level. In addition, organizations will assess performance of individual weather personnel, identify improvement areas and direct additional training as required. (T-2).
- 2.4.4. Provide TAFVER metrics for all supported locations to their parent MAJCOM and, upon request, to their supported unit commanders. (T-1).
- 2.4.5. Cross-feed any improved verification methods or tools developed to their parent MAJCOM. (T-1).

2.5. Organizations that produce WPs will:

- 2.5.1. Establish and maintain an OPVER program to measure WP performance analyze trends and implement training as required. (T-1).
- 2.5.2. Document all WP criteria that will be considered a “criteria event” (based on operational impacts as determined by the supported unit/activity) and provide to MAJCOMs for reference. (T-2).
- 2.5.3. Collect, analyze and report OPVER metrics according to [Chapter 5](#). (T-1).
- 2.5.4. Provide OPVER metrics to their parent MAJCOM and, upon request, to their supported unit commanders. (T-1).

2.5.5. Develop, implement, and document processes to use OPVER metrics to identify shortfalls in characterization products, approved techniques, internal processes, training, and certification. Document findings and corrective actions taken. (T-1).

2.5.6. Analyze OPVER metrics to include performance of individual weather personnel for internal improvement processes. (T-3).

2.5.7. Request MAJCOM technical assistance if needed to analyze and exploit results from OPVER metrics. (T-2).

2.5.8. Cross-feed any improved verification methods or tools developed to their parent MAJCOM. (T-1).

2.6. Organizations that produce MEFs will:

2.6.1. Establish and maintain an OPVER program to measure MEF performance, analyze trends and implement training as required. (T-1).

2.6.2. Document all MEF weather criteria that will be considered a “criteria event” (based on operational impacts) and provide to MAJCOMs for reference. (T-2).

2.6.3. Collect, analyze and report OPVER metrics according to [Chapter 5](#) and any additional guidance provided by MAJCOMs and/or chain of command. (T-1).

2.6.4. Provide OPVER metrics to their parent MAJCOM and, upon request, to their supported unit commanders. (T-1).

2.6.5. Develop, implement, and document processes to use OPVER metrics to identify shortfalls in characterization products, approved techniques, internal processes, training, and certification. Document findings and corrective actions taken. (T-1).

2.6.6. Analyze OPVER metrics to include performance of individual weather personnel for internal improvement processes. (T-3).

2.6.7. Request MAJCOM technical assistance if needed to analyze and exploit results from OPVER metrics. (T-2).

2.6.8. Cross-feed improved verification methods and tools developed to their parent MAJCOM. (T-1).

2.7. Organizations that produce Graphical Weather Depictions will:

2.7.1. Establish and maintain a GRAPHVER program to measure their product’s performance, analyze trends and implement training as required. (T-1).

2.7.2. Collect, analyze and report GRAPHVER metrics according to [Chapter 6](#) and any additional guidance provided by MAJCOMs and/or chain of command in supplements to this regulation. (T-1).

2.7.3. Provide GRAPHVER metrics to their parent MAJCOM and, upon request, to their supported unit commanders. (T-1).

2.7.4. Develop, implement and document internal processes to use GRAPHVER metrics to identify training issues, certification and operations shortfalls. Document findings and corrective actions taken. (T-1).

2.7.5. Analyze GRAPHVER metrics to include performance of individual weather personnel for internal improvement processes. **(T-3)**.

2.7.6. Request HQ USAF/A3W technical assistance if needed to analyze and exploit results from GRAPHVER metrics. **(T-2)**.

2.8. Organizations that produce Numerical Weather Model Depictions will:

2.8.1. Establish and maintain a MODVER program to measure their product's utility, analyze trends, establish benchmarks, and implement changes as required. **(T-1)**.

2.8.2. Collect, analyze and report MODVER metrics according to [Chapter 7](#) and any additional guidance provided by MAJCOMs and/or chain of command in supplements to this regulation. **(T-1)**.

2.8.3. Provide MODVER results to supported unit commanders, and their parent MAJCOM NLT the 20th of the following month. **(T-1)**.

2.8.4. Develop, implement and document internal processes to use MODVER metrics to identify model performance strengths/weaknesses, and operations shortfalls, to include standardized products and conditional verification (e.g., based on synoptic situations). Document findings and report them to the parent MAJCOM. **(T-1)**.

2.8.5. Make MODVER MOPs and MODVER assessments (i.e., consumable “forecaster-ready” interpretations of MOPs) readily available (e.g., online) to supported units and headquarters functional staffs. Coordinate the means of achieving this through their parent chain of command. **(T-1)**.

2.8.6. Maintain an active and documented unit-level program for evaluating and integrating new and appropriate verification metrics to support the AF fielding of combat acquisitions and new numerical weather modeling capabilities. At a minimum, apply and evaluate the feasibility and usefulness of non-standard verification MOPs listed in [Table 7.1](#).

2.8.7. Request HQ USAF/A3W technical assistance if needed to analyze and exploit results from MODVER MOPs. **(T-2)**.

Chapter 3

WEATHER WATCH, WARNING AND ADVISORY VERIFICATION

3.1. WARNVER Guidance and Procedures. Forecast Warnings and Watches are special notices of weather events or conditions of such intensity as to pose a hazard to life or property for which the supported organization/customer has documented protective posture or protective actions. Forecast Advisories are special notices of weather conditions that have potential to impact operations and safety. WARNVER uses objective measurements to quantify performance of WWA issuance and provides technical readiness insight. Forecast warnings, advisories and lightning watches will be verified according to AFMAN 15-129, Volume 1, *Air and Space Weather Operations-Characterization*. (T-1).

3.2. WARNVER Statistical Evaluation Methods

3.2.1. WARNVER will include the minimum MOPs as defined in [Table 3.1](#). (T-1).

3.2.2. MOPs will be calculated and reported for all WWA criteria individually, to include raw monthly data used for all calculations. (T-1).

3.2.3. Combined MOPs will be calculated by including all forecast warnings, forecast advisories, and lightning watches into a single metric. The raw number of WWAs must be totaled to create the overall average MOP. Do not use the average score for each MOP category when calculating the single overall average MOP. (T-1).

3.2.3.1. The combined warning MOP will be calculated by including only forecast warnings. (T1).

3.2.3.2. The combined forecast advisories MOP will be calculated by including all forecast advisories. (T-1).

3.2.3.3. The combined watch MOP will be calculated by including only lightning watches. (T1).

3.2.4. MAJCOMs and their subordinate organizations may develop additional MOPs and include them in the required monthly data.

3.2.5. WARNVER technical readiness metrics will include the minimum MOPs as defined in [Table 3.2](#). (T-1).

3.2.6. WARNVER standards are included in [Table 3.3](#).

Table 3.1. WARNVER MOPs.

MOP	Individual Event Calculation (for all events that occur within the specified verification distance)	Combined Event Calculation (for all events that occur within the specified verification distance)
Met DLT Percentage	The total number of forecast events that occurred and met the full DLT divided by the total number of events that occurred.	The total number of forecast WWA events that occurred and met the full desired lead-time divided by the total number of WWA events that occurred.
Positive Lead Time Percentage	The total number of forecast events with positive lead time divided by the total number of forecast events that occurred	The total number of WWA events with positive lead time divided by the total number of events that occurred
False Alarm Rate (FAR)	The number of issued WWAs minus the number of required WWAs divided by the number of issued WWAs.	The total number of issued WWAs minus the number of required WWAs divided by the total number of issued WWAs.
Mean Timing Error (MTE)	Timing error is the difference in time between the WWA forecast onset compared to the actual WWA time of occurrence. Calculate the absolute values of the timing errors per event, add them, and then calculate the average. For example, if four events had timing errors of -60 minutes, +60 minutes, -30 minutes and +30 minutes the total of the absolute values is 180 minutes, the MTE is 45 minutes.	Calculate the absolute values of the timing errors for all WWAs combined and then calculate a mean as described in the MTE for individual criteria to the left of this paragraph.
Required Not Issued (RNI) Percentage	The total number of WWAs that were required, but not issued, divided by the total number of events that met the event thresholds.	The total number of WWA events that were required, but not issued, divided by the total number of events that met the WWA thresholds.
Negative Lead Time Percentage	The total number of criteria WWAs issued with negative lead-time, divided by the total number of events that met the criteria thresholds. The total number of WWA events with negative lead-time divided by the total number of events that met the criteria thresholds.	The total number of criteria WWAs issued with negative lead-time, divided by the total number of events that met the criteria thresholds. The total number of WWA events with negative lead-time divided by the total number of events that met the criteria thresholds.

Table 3.2. WARNVER Technical Readiness MOPs.

Sub-threshold WWA (STW) Percentage	<p>Calculate the number of WWA events where the WWA criteria occurred but was one category less intense than the forecast intensity specified* divided by the total number of WWA events predicted for the intensity forecast. Note, this only applies to moderate or greater categories.</p> <p>* One WWA category less is the next reportable level below the WWA criteria. For example a 35-49-knot event is STW for a 50+ WWA. ½” hail is STW for a ¾” or greater hail warning. For WWAs concerning precipitation accumulation use one unit of measure below the warning threshold. For example, a heavy snow warning for 2” is STW for 1”.</p>	<p>Report separately for each criterion; also include a group with all forecast warnings combined and a group with all forecast advisories combined.</p>
Justified FAR	<p>The number of issued WWAs minus the number of required WWAs minus the number of WWAs that met 90% of the desired threshold (after issuance), divided by the number of issued WWAs.</p> <p>For lightning watches, use double the verification distance for Justified FAR, for example a 5-nautical mile (NM) WWA would be justified with strikes at or within 10 NM (that occur after issuance).</p>	<p>Report separately for each criterion; also include a group with all forecast warnings combined and a group with all forecast advisories combined. Note: Justified FARs are still counted as False Alarms.</p>

Table 3.3. WARNVER Standards.

MOP	Standard
Met DLT	Greater than, or equal to 75%
Positive Lead Time	Greater than, or equal to 90%
FAR	Less than, or equal to 40%
MTE	To be determined after data is collected and analyzed for a period not to exceed 18 months.
Negative Lead Time	Less than, or equal to 10%
NOTE: RNI WWA percentages are tracked internally and reported monthly according to Paragraph 2.3.5 if they occur.	

Chapter 4

TERMINAL AERODROME FORECAST VERIFICATION

4.1. TAFVER Guidance and Procedures. Timely, relevant, accurate and consistent TAFs provide meteorological information and form the foundation for mission execution, flight planning and command and control activities for a specific aerodrome complex. TAFVER uses objective measurements to quantify the accuracy of TAF production. The results of TAFVER provide information on forecast strengths, areas for improvement, recommended training areas, value added by the FITL and overall technical readiness. TAFVER is based on observed conditions throughout the valid period of the TAF.

4.2. TAFVER Statistical Evaluation Methods and MOPs

4.2.1. Evaluate the draft TAF generated by the model (if applicable) and the final FITL TAF. **(T-1).**

4.2.2. Evaluate TAFs using all available observations (Aerodrome routine meteorological report and Aerodrome special meteorological report). **(T-1).**

4.2.3. Use TAF code to measure performance for all groups that are forecast, becoming (BECMG), temporary (TEMPO) and from (FM). Determine if each group was correctly or incorrectly forecast for each hour. **(T-1).**

4.2.3.1. For a BECMG group to verify, forecast values can change up to 30 minutes before the start time and up to 29 minutes after the end time of the date/time group and must occur for at least 31 minutes each hour. **(T-1).**

4.2.3.2. For a FM group to verify, forecast values change at the time specified and must occur for at least 31 minutes each hour. **(T-1).**

4.2.3.3. For a TEMPO group to verify, forecast values change at the time specified and must occur at least once per hour, last less than one hour in each instance and in total, last less than half of the time indicated by TEMPO period. **(T-1).**

4.2.4. Compute TAFVER MOPs and technical readiness metrics, according to [Table 4.1](#) and [Table 4.2](#) for all groups that are forecast in the initial FITL TAF. The initial FITL TAF is valid at the issue time; amendments are not required to be scored. **(T-1).**

4.2.5. MAJCOMs and their subordinate organizations may develop additional TAFVER MOPs and technical readiness metrics as required and include them in monthly reports.

4.2.6. TAFVER Standards. TAFVER standards will be determined by HQ USAF/A3W in the future as reports of MOPs in this chapter are analyzed for a period of time not to exceed 18 months. **(T-2).** An Interim change to this publication may be issued when standards are ready for publication.

Table 4.1. TAFVER MOPs.

Criteria	Evaluate	Requirement	Hourly Score and Overall Percentage Correct
Ceiling (cig)	All specification and amendment criteria as documented on the Installation data page/Installation Weather support plan (or equivalent)	Verify within forecast categories as a correct forecast or an incorrect forecast for all groups.	<p>The hourly score is one point for a correct forecast and zero points for an incorrect forecast.</p> <p>The overall TAF cig percentage correct is the total number of points for correct forecasts (pcf) divided by the total number of available points (ap) multiplied by 100 $((pcf/ap)*100)$.</p>
Visibility (vis)	All specification and amendment criteria as documented on the Installation data page/Installation Weather support plan (or equivalent)	Verify within forecast categories as a correct forecast or an incorrect forecast for all groups.	<p>The hourly score is one point for a correct forecast and zero points for an incorrect forecast.</p> <p>The overall TAF vis percentage correct is the total number of points for correct forecasts (pcf) divided by the total number of available points (ap) multiplied by 100 $((pcf/ap)*100)$.</p>
Wind Speed	+ or – 9 knots	Verify all forecast groups where wind speeds are GTE than 6 knots. If the forecast is within 9 knots it is a correct forecast. For 10 knots or greater of error the forecast is incorrect.	<p>The hourly score is one point for a correct forecast and zero points for an incorrect forecast.</p> <p>The overall TAF Wind Speed percentage correct is the total number of points for correct forecasts (pcf) divided</p>

			by the total number of available points (ap) multiplied by 100 ((pcf/ap)*100).
Wind Direction	+ or – 50/30 degrees	Verify all forecast groups. For periods when winds are more than 6 knots but less than 15 knots, if the forecast direction is within 50 degrees, the forecast is correct. For periods when winds are greater than, or equal to, 15 knots, if the forecast direction is within 30 degrees the forecast is correct. When the forecast error is greater than these thresholds, the forecast is incorrect.	<p>The hourly score is one point for a correct forecast and zero points for an incorrect forecast.</p> <p>The overall TAF Wind Direction percentage correct is the total number of points for correct forecasts (pcf) divided by the total number of available points (ap) multiplied by 100 ((pcf/ap)*100).</p>
Wind Gusts	+ or – 10 knots of observed gusts	If gusts occur and are within 10 knots of the forecast criteria or no gusts are forecast and no gusts occur, it is counted as a correct forecast. For all cases where gusts are not forecast and gusts occur, no points are awarded.	<p>The hourly score is one point for a correct forecast and zero points for an incorrect forecast.</p> <p>The overall TAF Wind Gust percentage correct is the total number of points for correct forecasts (pcf) divided by the total number of available points (ap) multiplied by 100 ((pcf/ap)*100).</p>
Present Weather	Each phenomena separately, precipitation in liquid, freezing, or frozen, obscurations, and other. Intensity/proximity qualifiers are not mandatory for	<p>Verify all forecast groups using the Critical Success Index (CSI), which is correct forecast / (correct forecast + incorrect forecasts). Total score ranges from 1 to 0.</p> <p>An incorrect forecast is when a phenomena is forecast but not observed or not forecast</p>	<p>The hourly score is the hourly CSI, with a perfect forecast = to 1 point, and less than a perfect forecast a fraction of a point as defined by the CSI formula.</p> <p>The overall TAF</p>

	verification.	but was observed.	present weather score is the sum of the number of points (and fractions of a point) awarded each hour each for correct forecasts (pcf) divided by the total number of available points (ap) multiplied by 100 $((pcf/ap)*100)$.
Lowest Altimeter Setting	Measured for every forecast group (except TEMPO).	Verify within forecast categories as a correct forecast or an incorrect forecast as follows: If the lowest altimeter observed during a given hour was no more than .05 inches (ins) Hg lower than forecast during that hour it counts as a correct forecast. If the lowest altimeter observed during a given hour was more than .05 ins Hg lower than forecast during that hour it counts as an incorrect forecast.	<p>The hourly score is one point for a correct forecast and zero points for an incorrect forecast.</p> <p>The overall TAF lowest altimeter setting percentage correct is the total number of points for correct forecasts (pcf) divided by the total number of available points (ap) multiplied by 100 $((pcf/ap)*100)$.</p>
Combined TAF Accuracy	The overall TAF score using all available points earned divided by the possible available points for every hour in the TAF for all groups that were forecast, BECMG, TEMPO, and FM.	<p>Compute the sum of the total points correctly forecast (pcf) per group and divide by the sum of the total available points (ap) per group.</p> <p>$(BECMG\ pcf + TEMPO\ pcf + FM\ pcf) / (BECMG\ ap + TEMPO\ ap + FM\ ap)$</p>	

Table 4.2. TAFVER Technical Readiness Metrics.

Criteria	Requirement (BCMG, TEMPO and FM)	Hourly Score and Overall Percentage Correct (BCMG, TEMPO and FM)
Category cig Accuracy	As described in Table 4.1 for individual weather personnel. Identify individual skills and deficiencies and take actions as necessary.	As described in Table 4.1 for individual weather personnel.
Category vis Accuracy	As described in Table 4.1 for individual weather personnel. Identify individual skills and deficiencies and take actions as necessary.	As described in Table 4.1 for individual weather personnel.
Category cig bias	Number of total hours forecast for each cig category divided by the number of hours observed in each cig category.	Report scores by hour in the TAF and an overall score for all hours of the TAF.
Category vis bias	Number of total hours forecast for each vis category divided by the number of hours observed in each vis category.	Report scores by hour in the TAF and an overall score for all hours of the TAF.
Present Weather Accuracy	As described in Table 4.1 for individual weather personnel. Identify individual skills and deficiencies and take actions as necessary.	As described in Table 4.1 for individual weather personnel.
Present Weather Bias	Number of total hours forecast for each present weather event divided by the number of hours observed in each present weather category.	Report scores by hour in the TAF and an overall score for all hours of the TAF.
FITL Value Added	Compute TAFVER MOPs according to Table 4.1 for the model produced TAF (if applicable). Subtract the model produced TAF MOPs from the FITL TAF MOPs to determine the FITL value added.	Report scores by hour in the TAF for each MOP in Table 4.1 and include an overall score for all hours of the TAF.

Chapter 5

OPERATIONAL VERIFICATION

5.1. OPVER Guidance and Procedures. Timely, relevant, accurate and consistent WPs/MEFs directly impact decision superiority by enhancing predictive battlespace awareness and by enabling friendly forces to anticipate and exploit the battlespace environment. OPVER will not be limited to MEFs and will include WPs provided hours or even days in advance that impact operational planning. **(T-1).** OPVER will use objective measurements (when available) to quantify the accuracy of weather products and processes used to support operations; subjective data is a backup and only used in the absence of objective data. **(T-1).**

5.1.1. OPVER assesses the relevance of WPs to mission planning and execution, and identifies relationships between forecast criteria for all operations and observed weather conditions. The results of the OPVER process provide information on operational impacts and forecast performance spanning the entire spectrum of the supported organization's operations. Weather personnel must understand how atmospheric and space weather effects can influence the integration, enable mitigation and maintain a relevant OPVER program.

5.1.2. OPVER focuses on identification of weather and space environmental criteria that are significant to operations and tactics. In addition, OPVER procedures attempt to track how often operators take mitigation actions based on MEFs and WPs and how often those actions were necessary. This provides insight into our performance to communicate impacts to operators and encourage them to take actions. Some missions or sorties may not have the luxury to reschedule or modify based on weather, but many will and tracking this information is as important as determining if we had the forecast correct. Elements assessed may include meteorological, solar, ionospheric, oceanographic or other significant geophysical phenomena.

5.2. OPVER Scope. OPVER will be computed for all MEFs and WPs that are forecast to reach or exceed a forecast criteria condition and for all occurrences of forecast criteria conditions. **(T2).**

5.3. OPVER Statistical Evaluation Methods and MOPs.

5.3.1. Statistical Evaluation Methods.

5.3.1.1. Define and document WP and MEF mission impacting weather criteria to include at a minimum: takeoff, route/operating area, landing and divert criteria for every supported aviation platform and mission. **(T-2).**

5.3.1.1.1. Mission-impacting criteria should include but are not limited to: platform limitations, pilot categories, training limitations, airfield minima, and tactical requirements.

5.3.1.1.2. Include all Mission Support Group, Maintenance Group, Medical Group, and Tenant organization required mission-impacting weather criteria. **(T-2).**

5.3.1.1.3. For transient aircraft and/or unknown mission limitations, organizations will use standard IFR/Visual Flight Rule thresholds when computing OPVER. **(T-2).**

5.3.1.2. Identify and document all supported organizations and operational users on and off of the airfield to provide structured feedback on WPs and MEFs used to support their mission(s). Documentation must be signed off at the Group or equivalent level and may be but is not limited to a weather support document/plan. **(T-2).**

5.3.1.3. Conduct post-mission analysis for all FITL WPs and MEFs. This excludes products covered by GRAPHVER procedures. If a WP is subject to a Meteorological Watch (METWATCH) and amended, it must be verified for all events that are forecast or observed to meet a forecast criteria. Where available, use objective data vice subjective data to verify WP or MEF element accuracy. Objective verification data may include, but is not limited to: radar-derived parameters, direct pilot or supported commander feedback, surface weather observations, lightning detection data, formal pilot debriefs and pilot reports (PIREP) from mission aircraft. It is essential that weather organizations producing WPs and MEFs complete the process by obtaining post-mission analysis data. **(T-2).**

5.3.1.4. Conduct subjective post-mission analysis when objective data is unavailable or when subjective data would assist in determining whether elements of every WP or every MEF verified. Subjective verification data includes sources deemed credible by unit leadership and may include, but is not limited to: satellite imagery interpretation, PIREPs from other aircraft in the vicinity of the mission area and other credible weather reports. **(T-2).**

5.3.1.5. Regardless of the mission outcome (proceed as is/cancel/change), OPVER will be conducted for all MEFs requiring verification according to **Paragraph 5.3**. Sometimes, several MEFs may be necessary to complete a mission. For example, if a MEF is presented, and the mission director changes the mission time to correspond with conditions that are more favorable, then a new MEF may be required or an update to the previous one to match the mission changes. These count as two MEFs, both requiring verification. **(T-3).**

5.3.1.6. To compute OPVER metrics, weather organizations will:

5.3.1.6.1. Use the OPVER computation grid in **Table 5.1** to standardize the collection and analysis of WP and MEF data. WPs will be verified at the beginning and ending of the valid period, and hourly through the first six hours, there after products will be verified every three hours. **(T-1).** For example, a WP valid for 6 hours will be verified at the beginning of the valid time, and hourly thereafter. A WP valid for 14 hours will be verified at the beginning of the valid period, hourly through the first 6 hours, at the 9, 12 and 14 hour points.

5.3.1.6.2. Record WP and MEF verification totals separately using **Table 5.1**. **(T-1).**

5.3.1.6.2.1. Block A (mandatory): “Criteria Event WP, Criteria Event Observed.” The total in block A is the number of correct WP Criteria Event WP forecasts.

5.3.1.6.2.2. Block B (mandatory): “No Criteria Event WP, Criteria Event Observed”. The total in block B is the number of incorrect WP No Criteria Event WP forecasts.

5.3.1.6.2.3. Block C (mandatory): “Criteria Event WP, No Criteria Event Observed”. The total in block C is the number of incorrect WP Criteria Event WP forecasts.

5.3.1.6.2.4. Block D (optional for manual calculations): “No Criteria Event WP, No Criteria Event Observed”. The total in block D is the number of correct WP No Criteria Event WP forecasts.

5.3.1.6.3. If available, record the monthly total of MEFs and WPs that resulted in a mission change (MEFC / WPC) to avoid forecast Criteria. **(T-2)**. For example, if 25 missions were modified based on MEFs to avoid a significant snow storm, a high wind event and a thunderstorm event, ensure that number is recorded as 25 MEFCs.

5.3.1.6.4. Compute WP and MEF metrics according to [Table 5.2](#) and [Table 5.3](#). **(T-1)**.

Table 5.1. WP/MEF Grid.

		FORECAST Conditions (MEF/WPs)		
		Criteria Event WPs/MEFs	No Criteria Event WPs/MEFs	Totals
OBSERVED Conditions	Criteria Event OBSERVED	A	B (Miss)	Total Criteria Event OBSERVED: A + B
	No Criteria Event OBSERVED	C (False Alarm)	D	Total No Criteria Event OBSERVED: C + D
	Totals	Total Criteria Event WPs/MEF: A + C	Total No Criteria Event WPs/MEFs: B + D	Total WPs/MEFs: A + B + C + D

Table 5.2. WP/MEF Metrics.

Performance Metric	Formula	Description
Accuracy (Optional)	$(A+D)/(A+B+C+D) \times 100\%$	Accuracy indicates the percentage of accurate WPs/MEFs compared to all WPs/MEFs issued. CAUTION: Do not use this metric alone to judge the overall performance of the WP program due to the naturally occurring high percentage of "No Criteria Event WPs, No Criteria Event Observed" outcomes at many operating locations. Use "Criteria Event Accuracy" and "No Criteria Event Accuracy" metrics to shed light on problem areas.
Criteria Event Accuracy (Mandatory)	$(A/(A+C)) \times 100\%$	Criteria Event Accuracy indicates percentage of Criteria Event WP/MEFs that verified correctly. This tells leadership how often a forecast for mission impacting weather verified.
No Criteria Event Accuracy (Optional)	$(D/(B+D)) \times 100\%$	No Criteria Event Accuracy indicates percentage of no operational criteria WP/MEF forecasts that were correctly made. This tells leadership how often a forecast for non-mission impacting weather verified.
Mitigation Rate (Optional)	$((MEFC + WPC) / (\text{criteria events forecast for MEFs/WPs})) \times 100$	Take the monthly total of MEFCs and WPCs (MEFs and WPs that resulted in mission changes) and divide by the monthly total of WP and MEF criteria event forecasts. Multiply by 100 to determine the mitigation rate, or percentage of time operators took action on criteria event forecasts for WPs/MEFs. Generally speaking the higher the mitigation rate the more successful the MEF/WP program is. The most successful outcome of a mission impacting forecast is when operators accept the input and change their mission profiles to mitigate the risk. There will be instances where operators cannot change missions and must try to accomplish the sortie or mission despite a forecast. Consider tracking those situations separately.

Table 5.3. WP/MEF Technical Health Metrics.

Performance Metric	Formula	Description
Criteria Event Bias (Mandatory)	$(A+C)/(A+B)$	Criteria Event Bias reveals whether mission impacting events were either over or under forecast. Criteria Event WP Bias > 1 means Criteria Event WPs were over forecast. Criteria Event WP Bias < 1 means Criteria Event WPs were under forecast. For example, a Criteria Event Bias of 2 means mission impacting events were forecast 200% more than they occurred, a Criteria Event of 0.5 means mission impacting events were under forecast 50% of the time. It is important to compare Criteria Event Bias to the Criteria Event accuracy. An ideal balance would show the capability to predict mission impacting events without a high level of over forecasting.
No Criteria Event Bias (Optional)	$(B+D)/(C+D)$	“No Criteria Event WP Bias” reveals whether non-mission impacting weather forecasts events were either over or under forecast. “No Criteria Event Bias” > 1 means non-mission impacting weather forecasts were over forecast. “No Criteria Event Bias” < 1 means non-mission impacting weather forecasts were under forecast. For example, a No Criteria Event Bias of 2 means non-mission impacting weather forecasts were forecast 200% more than they occurred, a No Criteria Event of 0.5 means non-mission impacting weather forecasts events were under forecast 50% of the time. It is important to compare this metric with the No Criteria Event Accuracy.

Chapter 6

GRAPHICAL WEATHER DEPICTION VERIFICATION

6.1. GRAPHVER Guidance and Procedures. Timely, relevant, accurate and consistent graphical weather depictions provide meteorological information for mission execution, flight planning and command and control activities for large geographic areas. GRAPHVER uses a significant amount of subjective measurements and limited objective measurements to quantify the accuracy and performance of graphical weather depictions. Subjective measurements are allowed according to AFMAN 15-129, Volume 1, to ensure accurate and repeatable corrective actions. The results of GRAPHVER provide information on forecast strengths, areas for improvement, recommended training areas, and overall technical readiness.

6.2. GRAPHVER is based on observed conditions or derived evaluations based on reliable information throughout the valid period of the graphics product in addition to subjective verification data points. Manual verification is done for a representative sample as outlined in this instruction. At a minimum, all graphic weather products subject to METWATCH and are amended will be verified. (T-2). This includes aviation hazard depictions for turbulence, icing, and thunderstorms as well as any graphical MOAF products. (T-2).

6.3. Objectively grade turbulence and icing product sets. Aviation hazards will be evaluated for product accuracy in correctly specifying areas of moderate or greater criteria and provided an “impact” grade to determine overall forecast capabilities. (T-2).

6.3.1. Units will verify a minimum of two full upper and lower level turbulence and icing product sets created based on the 00Z and 12Z model forecast data for each forecast panel from 03-33 hours daily. (T-2).

6.3.1.1. Construct an evaluation set for turbulence and icing forecasts by creating a composite image containing PIREPS, air reports, or other significant information collected for the cardinal hour of the valid period of the product (e.g., all PIREPS from 1501-1559Z) and a meteorological satellite image with a date/time group nearest the midpoint of the cardinal hour of the valid period of the product (e.g., 1530Z) of the original icing or turbulence forecast product. Evaluation of updated or amended product sets is at unit’s discretion. (T-2).

6.3.1.2. Verification will be completed by superimposing a 90-NM x 90-NM grid box set over forecast areas (FAs) and all observed areas. (T-1). Area coverage of the forecast determines how the grid boxes will be evaluated and scored. (T-1).

6.3.1.2.1. Grid boxes with greater than or equal to 50% or more of their surface area covered by forecast criteria (moderate or greater turbulence or icing) are considered to be FAs for purposes of evaluation. Since the evaluation grid box scaling corresponds with the amendment criteria, any report within the grid box will validate a forecast even though the actual report may be outside the area drawn on the forecast product. (T-1).

6.3.1.2.2. Grid boxes with less than 50% of the surface area covered by a forecast criteria must contain reports in order to be evaluated as a FA. Evaluation grid boxes

less than 10% covered by a forecast criteria are considered forecast hazard free for purposes of evaluation. **(T-1)**.

6.3.1.2.3. FAs with corresponding reports of moderate or greater turbulence or icing are impact verified boxes (V). Track and report metrics for moderate criteria, and severe combined with occasional severe separately. **(T-1)**.

6.3.1.2.4. Grid boxes containing reports of moderate or greater turbulence or icing outside of forecast areas (forecast light with reported severe or forecast none with reported moderate) are impact required, not forecast (RNF) boxes. **(T-1)**.

6.3.1.2.5. Grid boxes containing sub-threshold reports (forecast moderate with reported light) are non-verified boxes. **(T-1)**.

6.3.1.3. Determine the forecast reliability of turbulence and icing products by dividing the total number of verified boxes by the sum of the number FA boxes and the number required not forecast boxes; $V / (FA + RNF)$. **(T-1)**.

6.3.1.4. Determine the required not forecast rate by dividing the total number of required not forecast boxes by the total number of all grid boxes. **(T-1)**.

6.4. Objectively grade the thunderstorm product sets. Aviation thunderstorm hazard forecasts will be evaluated for product accuracy in correctly specifying the occurrence of thunderstorm activity in the respective FAs. **(T-1)**. Areal extent of thunderstorm forecasts can be evaluated based on sensed lightning and RADAR returns (if available), but instantaneous coverage forecast specifications (Isolated, Few, etc.) cannot reliably be evaluated after the fact. A high lightning strike count does not directly correlate to numerous thunderstorm cells, nor does a low strike count indicate isolated coverage within a FA. Global RADAR coverage is insufficient to employ these sensors as an evaluation source both due to coverage gaps and the acknowledgement that high reflectivity does not always indicate the presence of lightning. With these technical limitations in mind, evaluation of thunderstorm prognoses will focus on the ability to accurately specify the areal extent of thunderstorms during the valid period of the forecast. **(T-1)**.

6.4.1. Units will verify a minimum of two full thunderstorm product sets created based on the 00Z and 12Z model forecast data for each forecast panel from 03-33 hours, daily. **(T-1)**.

6.4.1.1. Construct an evaluation set for thunderstorm forecasts by creating a composite image containing lightning strike data for the 3-hour valid period of the product (09Z chart is valid from 09-12Z), a composite reflectivity mosaic (if available) and a meteorological satellite image valid for the point in time nearest the midpoint of the forecast product (e.g. Use the 1930Z imagery for a thunderstorm prognosis valid 1800-2100Z) combined with the original thunderstorm forecast product. Evaluation of amended product sets is at the unit's discretion. **(T-1)**.

6.4.1.2. Verification will be completed by superimposing a 90NM x 90NM grid box set over FAs and all observed areas. **(T-1)**. Area coverage of the grid box determines how evaluation boxes will be scored. **(T-1)**.

6.4.1.2.1. Grid boxes with greater than or equal to 50% or more of their surface area covered by forecast criteria (thunderstorm area) are considered to be FAs for purposes of evaluation. Since the evaluation grid box scaling corresponds with the amendment

criteria, any report within the grid box will validate a forecast even though the actual report may be outside the area drawn on the forecast product. **(T-1)**.

6.4.1.2.2. Grid boxes with less than 50% of the surface area covered by a forecast criteria must contain reports in order to be evaluated as a FA. Evaluation grid boxes less than 10% covered by a forecast criteria are considered forecast hazard free for purposes of evaluation. **(T-1)**.

6.4.1.2.3. FAs with corresponding lightning strikes and thunderstorm activity, detected on the composite reflectivity mosaic (if available) and meteorological satellite imagery, are impact verified boxes (V). **(T-1)**.

6.4.1.2.4. FAs with no observed lightning strikes or thunderstorm activity are considered non-verified boxes. **(T-1)**.

6.4.1.2.5. Grid boxes containing lightning strikes or thunderstorm signatures on the composite reflectivity mosaic or meteorological satellite imagery, that are outside of FAs, are impact RNF boxes. **(T-1)**.

6.4.1.3. Determine the forecast reliability of thunderstorm products by dividing the total number of verified boxes (V) by the sum of the number FA boxes and the number RNF boxes; $V / (FA + RNF)$. **(T-1)**.

6.4.1.4. Determine the required not forecast rate by dividing the total number of RNF boxes by the total number of all grid boxes. **(T-1)**.

Chapter 7

NUMERICAL WEATHER MODEL VERIFICATION (MODVER)

7.1. MODVER Guidance and Procedures. Timely, relevant, accurate and consistent numerical weather model predictions provide a significant part of the foundation for meteorologists to build mission execution, flight planning and command and control activities for large and small scale geographic areas. MODVER uses a significant amount of objective measurements to quantify the accuracy, skill, value, and performance of weather models. MODVER results provide information on weather model strengths and weaknesses, help scientists and decision makers identify areas for improvement, and support strategic partnerships through the application and reporting of regional and international MOPs. MODVER metrics inform decisions, but alone do not make the decision an empowered human analyzes metrics in relationship to the overall question being asked and makes an informed decision based on data in relevant decision focus areas.

7.1.1. MODVER is based on forecast weather model conditions and those observed weather conditions throughout the model valid period. Units that create weather model predictions will use statistical analysis to create MOPs and standards for their models. **(T-1)**. The measures of performance will be created to compare models to standards of performance, in **Table 7.2**, **Table 7.3** and **Table 7.4**, as applicable. **(T-1)**.

7.1.2. AF-led modeling efforts with international and or interagency partners may require specific measures of performance not listed in the tables below. HQ USAF/A3W will direct agencies to report any required MOPs at the level of detail and frequency that is required through their applicable MAJCOM.

7.2. Statistical Evaluation Methods and MOPs

7.2.1. Evaluate forecast model direct output and derived forecast variables using the appropriate observations sources, to include station-based, gridded analysis data, and/or remotely sensed (e.g. satellite, radar, etc.). **(T-1)**.

7.2.2. Establish and employ automated and/or manual quality control procedures on observational data prior to use in MODVER and scientifically determine if the observation source is adequate for the purpose of MODVER. **(T-1)**.

7.2.3. Evaluate multiple types of MOPs. **Table 7.2** lists the minimum MOPs to be used for all numerical weather model (NWM) output. **(T-1)**. However, these should almost always be supplemented by additional MOPs appropriate for the NWM characteristics and phenomenon being predicted. Suggested supplemental MOPs may include but are not limited to those listed in **Table 7.3** (for deterministic NWM output) and **Table 7.4** (for stochastic NWM output).

7.2.3.1. Employ scale and physically appropriate MOPs. Traditional approaches to MODVER (i.e. precise matching of a single observation point to a single forecast point, contingency table, root mean square error metrics, etc.) do not adequately assess the quality or value of high resolution models (i.e., model grid-spacing less than 5 km, a.k.a. “kilometer scale”) and may mask poor representation problems with coarse resolution models.

7.2.3.2. Evaluate probabilistic model forecasts (e.g. ensemble output) with appropriate MOPs. **(T-2)**.

7.2.3.3. Employ spatially or object aware MOPs for forecast model direct output and derived variables, as appropriate. **(T-2)**. Neighborhood and spatially or object-aware MOPs more closely examine a model's performance with regards to storm structure and organization, key elements in an assessment of model's ability to resolve features at all appropriate scales (i.e., convective to synoptic).

Table 7.1. Definition of Variables.

Variable	Description
n	Number of forecast/observation pairs
m	Model forecast value
o	Observed value
c	Climatological value (i.e., long-term mean)
y	Stochastic forecast probability of event occurrence
p	Observed event occurrence (0 or 1)
w_j	Weighting assigned to the j th forecast parameter
SS_j	Skill Score for the j th forecast parameter
ref	Reference forecast (often output from another NWP model)
k	An event occurrence threshold
K	Total number of event occurrence thresholds
O_i	Used in Fractional Skill Score. For a given location, i , it is the fraction of the surrounding area (e.g., a 10-km ring) that is observed to meet the criteria threshold (e.g., thunderstorm)
M_i	Used in Fractional Skill Score. For a given location, i , it is the fraction of the surrounding area (e.g., a 10-km ring) that is predicted to meet the criteria threshold (e.g., thunderstorm).
$F_i^f(x)$	Cumulative distribution function of the stochastic forecast probability for the i th forecast case
$F_i^o(x)$	Cumulative distribution function of the observation for the i th forecast case (normally a single stepwise function from 0 to 1 at the observed value)
A, B, C, and D	Same as in Table 5.1 .
A_{MODE} , B_{MODE} , and C_{MODE}	Same as “A”, “B”, and “C” from Table 5.1 , but based on a forecast field matching procedure from the Method for Object-Based Diagnostic Evaluation (MODE). A_{MODE} refers to an observed area matched to a corresponding FA, B_{MODE} refers to an observed area not matched to a corresponding forecast are, and C_{MODE} refers to a forecast area not matched to a corresponding observed area.

Table 7.2. Required Minimum MOPs.

Performance Metric	Example Equations	Description
Mean Square Error (MSE)	$\frac{1}{n} \sum (m - o)^2$	Primary metric to measure model accuracy. Observed values should be from station data as opposed to model analysis. Required parameters are temperature at 500 hecto Pascals (hPa) and 2 meters, wind direction and speed at 500 hPa and 2 meters, and sea-level pressure.
Bias	$\frac{1}{n} \sum (m - o)$	The average error, with sign, of the predicted and observed value. Observed values should be from station data as opposed to model analysis. Required parameters are temperature at 500 hPa and 2 meters, wind direction and speed at 500 hPa and 2 meters, and sea-level pressure.
Anomaly Correlation	$\frac{\sum \{[(m - \bar{c}) - \overline{(m - \bar{c})}][(o - \bar{c}) - \overline{(o - \bar{c})}]\}}{\sqrt{\sum [(m - \bar{c}) - \overline{(m - \bar{c})}]^2 \sum [(o - \bar{c}) - \overline{(o - \bar{c})}]^2}}$	Commonly used metric at modeling centers to assess model accuracy. Verifies NWM output against a model analysis (usually from the same model) rather than against observations. Also has the benefit of not rewarding a lack of variability in the model forecast, which can occur with MSE. Required minimum parameter is 500 hPa geopotential height, but other variables can be used as well.

Table 7.3. Optional MOPs for Determinist NWM Output

Performance Metric	Example Equations	Description
Skill Score	$\left(1 - \frac{MSE}{MSE_{ref}}\right)$	Useful for comparing performance of two models.
Generalized Operations Index	$\sqrt{1 - \frac{1}{\sum w_j} \left(\sum w_j SS_j\right)}$	Useful for consolidating the Skill Score for multiple forecast parameters into a single index. The skill score for each parameter, j , is assigned a weight based on its importance to the overall index.
Root Mean Square Error (RMSE)	$\sqrt{\frac{1}{n} \sum (m - o)^2}$	Similar to MSE, another common measure of NWM accuracy.
Probability of Detection (POD)	$(A/(A+B)) \times 100\%$	For measuring NWM ability to predict occurrence of a specific event, such as 25-kt surface winds. POD measures the percentage of all event occurrences that were predicted by the model.
FAR	$(C/(A+C)) \times 100\%$	For measuring NWM ability to predict occurrence of a specific event, such as 25-kt surface winds. FAR measures the percentage of all predicted event occurrences in which the event did not occur.
CSI	$A/(A+B+C)$	For measuring NWM ability to predict occurrence of a specific event, such as 25-kt surface winds. CSI takes into account both POD and FAR, and is therefore a more balanced single metric. Tends to excessively penalize predictions of rare events, so is better used when threshold is routinely observed. Values range from 0 to 1 (perfect forecast).
Root Mean Square Factor (RMSF)	$\exp \left(\sqrt{\frac{1}{n} \sum \left[\log \left(\frac{m}{o} \right) \right]^2} \right)$	Similar to MSE, another common measure of NWM accuracy. A perfect forecast is 1, with values greater than 1 indicating greater model error. RMSF penalizes an under-forecast more than an over-forecast, and is therefore useful for sensible weather (e.g., wind gusts), when an under-forecast may be particularly undesirable.
Fractions Skill Score	$1 - \frac{\sum (O_i - M_i)^2}{\sum O_i^2 + \sum M_i^2}$	This is a spatially-aware MOP useful for evaluating NWM ability to predict location and spatial extent of an event, such as thunderstorms, high winds, dust storms, or tornados. Values range from 1

		(perfectly-matched location) to 0 (complete mismatch).
MODE	$A_{MODE}/(A_{MODE}+B_{MODE}+C_{MODE})$	This is an object-aware MOP best suited for verification of highly localized, irregular fields, such as thunderstorms, high winds, dust storms, or tornados. There are many variations of the MODE MOP, but all entail some sort of procedure to determine if observed fields of the phenomenon are or aren't "matched" by a corresponding forecast field from the NWM output. A contingency table is built similar to that shown in Table 5.2 ; but, it is based on hits, misses, and false alarms from the matching procedure. Once this table is built, the MODE MOP can be expressed using the equation shown; the same as that used to compute CSI. The procedure may also examine attributes of the fields, such as intensity, area, axis angle, aspect ratio, or curvature, and incorporate these into the final statistics.

Table 7.4. Optional MOPs for Stochastic NWM Output

Performance Metric	Example Equations	Description
Brier Score (BS)	$\frac{1}{n} \sum (y - o)^2$	For measuring ability of stochastic NWM output to predict occurrence of a specific event, such as 25kt surface winds.
Brier Skill Score	$1 - \frac{BS}{BS_{ref}}$	Useful for comparing performance of two stochastic models to predict occurrence of a specific event.
Rank Probability Score (RPS)	$\frac{1}{K-1} \sum_k BS_k$	Similar to BS, except it consolidates stochastic model performance at multiple event thresholds into a single metric.
Rank Probability Skill Score	$\left(1 - \frac{RPS}{RPS_{ref}}\right)$	Compares RPS for two different stochastic models.
Continuous Ranked Probability Score (CRPS)	$\frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{+\infty} (F_i^f(x) - F_i^o(x))^2 dx$	A thorough metric that measures stochastic model accuracy at all forecast thresholds for a given parameter.
Continuous Ranked Probability Skill Score	$\left(1 - \frac{\overline{RPS}}{\overline{RPS_{ref}}}\right)$	Compares CRPS for two different stochastic models.

7.2.3.4. At a minimum and to maintain a consistent baseline, the MOPs list in [Table 7.2](#) will be computed for the indicated parameters at forecast hours of 6, 12, 24, 48, 72, and

120 for a global model. **(T-1)**. At a minimum, higher-resolution limited-area models should be verified at these same forecast hours excluding the hours beyond the length of the forecast run (i.e., a limited-area model that extends to 48 hours should, at a minimum, be verified at forecast hours of 6, 12, 24, and 48.) Additional variables, levels, and forecast hours should be included as appropriate for the standardized MOPs listed in Table 7.2 and any supplemental MOPs. Variables may be direct from the model output (e.g., temperature, moisture, wind, etc.) or derived (e.g., visibility, max wind gust, etc.).

7.2.3.5. When selecting supplemental MOPs, make selections appropriate for the physical scale of the primary phenomena being predicted, as well as the intended purpose of the NWM output. Many traditional approaches to MODVER (i.e. precise matching of a single observation point to a single forecast point, contingency table, root mean square error metrics, etc.) inherently reward coarse resolution output's lack of small scale features, which often punish high-resolution models if not located and timed precisely. In some machine-to-machine applications of NWM output (e.g., flight level winds), a lack of small-scale features may not be an issue, and traditional metrics may be suitable. On the other hand, if the NWM output is to be used for predicting severe weather or other small-scale features, supplemental MOPs are necessary to adequately assess NWM performance. This is especially true when the NWM output is used as a tool for assessing the threat of extreme or small-scale events.

7.2.3.6. If a primary purpose of the NWM output is to predict "events" (thunderstorms, high winds, dust storms, tornados, etc.), employ spatially- or object-aware MOPs such as the Fractions Skill Score or MODE listed in [Table 7.3. \(T-1\)](#). These MOPs more closely examine a model's performance with regards to storm structure and organization, key elements in its ability to resolve features at all appropriate scales (i.e., convective to synoptic).

7.2.3.7. Although [Table 7.4](#) lists commonly used MOPs for stochastic NWM output, it is also possible to apply deterministic MOPs from [Table 7.2](#) to stochastic output, which might be desirable when assessing the added value of an ensemble. This is especially useful for NWM severe weather predictions using spatially- or object-aware MOPs. Deterministic verification of an ensemble can be performed by using the ensemble mean or a certain ensemble probability threshold as a deterministic forecast.

7.2.4. Subjective MOPs

7.2.4.1. The objective MOPs described in [Paragraph 7.2.3](#) are repeatable, require little to no interpretation, and are generally preferable for routine NWM performance assessments. However, in some instances, subjective verification is useful to compliment objective MOPs in order to capture aspects of model performance not easily defined by an equation.

7.2.4.2. Subjective verification relies on qualitative interpretation of NWM products and is based on thorough knowledge of meteorology and modelling and on experience gained in operational forecasting. It may be useful for evaluating derived and other unique meteorological parameters for which there is limited verification data. For instance, subjective verification may turn to non-traditional observations, satellite data, derived sounding products, or other data sources to qualitatively evaluate NWM predictions.

7.2.4.3. Subjective verification is best used when comparing two or more different NWM outputs. For example, a valid subjective MOP can be attained by having a group of experienced weather personnel use two different NWM outputs for multiple days, and having each individual subjectively grade the NWM output on a scale from 1 to 10 for how useful it was to the forecast process.

7.2.4.4. Subjective verification is far more useful when paired with objective MOPs to the extent possible. It also serves to independently corroborate and evaluate traditional and newer non-traditional objective methods to ensure verification procedures are performing as desired.

7.2.4.5. When performing subjective verification, units will maintain adequate documentation of the evaluation, methodology, and justification. **(T-1)**. Since subjective verification does not always entail rigidly-defined evaluation criteria, documentation is the only way to ensure key elements of the results are not misinterpreted over time.

MARK C. NOWLAND, Lt Gen, USAF
Deputy Chief of Staff, Operations

Attachment 1

GLOSSARY OF REFERENCES AND SUPPORTING INFORMATION

References

AFI 10-201, *Force Readiness Reporting*, 3 Mar 2016

AFPD 15-1, *Weather Operations*, 12 November 2015

AFI 15-128, *Air Force Weather Roles and Responsibilities*, 7 February 2011

AFI 33-360, *Publications and Forms Management*, 01 December, 2015

AFMAN 15-129, Volume 1, *Air and Space Weather Operations-Characterization*, 6 December 2011

AFMAN 33-363, *Management of Records*, 1 March 2008

Prescribed Forms

None

Adopted Forms

DD Form 175-1, *Flight Weather Brief*

AF Form 847, *Recommendation for Change of Publication*

Abbreviations and Acronyms

AF—Air Force

AFI—Air Force Instruction

AFMAN—Air Force Manual

AFPD—Air Force Policy Directive

AP—Available Points

BS—Brier Score

BECMG—Becoming

CIG—Ceiling

CRPS—Continuous Ranked Probability Score

CSI—Critical Success Index

DLT—Desired Lead Time

DZ—Drop Zone

FA—Forecast Area

FAR—False Alarm Rate

FITL—Forecaster-in-the-Loop

FM—From

GRAPHVER—Graphical Weather Depiction Verification

hPa—hecto-Pascal

IFR—Instrument Flight Rules

INS—Inches

LZ—Landing Zone

MAJCOM—Major Command

MEF—Mission Execution Forecast

MEFC—Mission Execution Forecasts that resulted in mission Changes

METWATCH—Meteorological Watch

MOAF—Military Operations Area Forecast

MODE—Method for Object-Based Diagnostic Evaluation

MODVER—Model Prediction Verification

MOP—Measures of Performance

MSE—Mean Square Error

MTE—Mean Timing Error

NLT—No Later Than

NM—Nautical Mile

NWM—Numerical Weather Model

OPR—Office of Primary Responsibility

OPVER—Operational Verification

PCF—Points for Correct Forecast

PIREP—Pilot Report

POD—Probability of Detection

RMSE—Root Mean Square Error

RMSF—Root Mean Square Factor

RNF—Required, Not Forecast

RPS—Rank Probability Score

STW—Sub-Threshold Weather Watch, Warning, and Advisory

TAF—Terminal Aerodrome Forecast

TAFVER—Terminal Aerodrome Forecast Verification

TEMPO—Temporary

USAF—United States Air Force

VIS—Visibility

WARNVER—Weather Watch, Warning, and Advisory Verification

WWA—Weather Watch, Warning, and Advisory

WP—Weather Product

WPC—Weather Product Changes